



HAL
open science

RNA inverse folding can be solved in linear time for structures without isolated stacks or base pairs

Théo Boury, Laurent Bulteau, Yann Ponty

► **To cite this version:**

Théo Boury, Laurent Bulteau, Yann Ponty. RNA inverse folding can be solved in linear time for structures without isolated stacks or base pairs. SEQBIM 2024 - Journées du groupe de travail Séquences en Bioinformatique, Nov 2024, Rennes, France. hal-04773768

HAL Id: hal-04773768

<https://hal.science/hal-04773768v1>

Submitted on 13 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

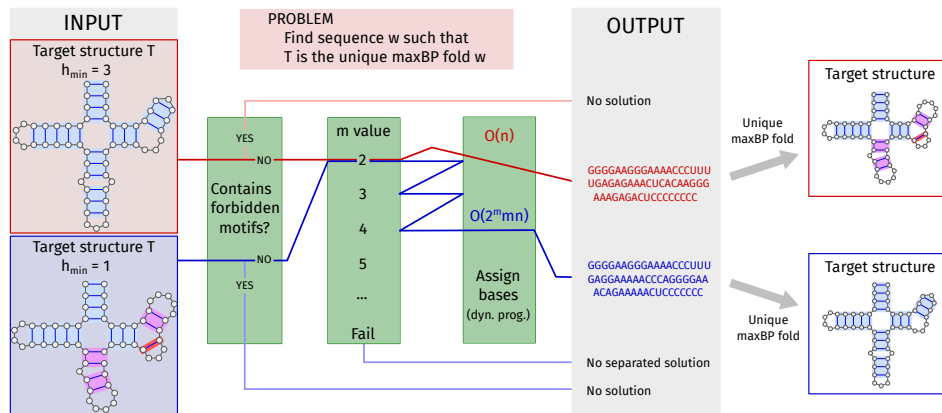
Abstract

RNA inverse folding can be solved in linear time for structures without isolated stacks or base pairs

Théo Boury¹, Yann Ponty^{1*}, Laurent Bulteau^{2*}¹Laboratoire d'Informatique de l'Ecole Polytechnique (LIX;UMR 7161), IP Paris, France²Laboratoire d'Informatique Gaspard-Monge (LIGM;UMR 8049), CNRS, Université Gustave Eiffel, France

*Corresponding authors: yann.ponty@lix.polytechnique.fr, laurent.bulteau@univ-eiffel.fr

Abstract



Inverse folding (IF) is a classic instance of negative RNA design, and consists in finding a sequence that uniquely folds into a target secondary structure with respect to free-energy minimization. Even when base pairs uniformly contribute -1 to the energy, and energy minimization coincides with maximizing the number of base pairs, the (exponential) numbers of potential alternatives over a candidate sequence makes IF a challenging problem to tackle. Accordingly, IF is generally believed to be NP-hard, as formally demonstrated under mild constraints by Bonnet *et al.* [1].

In our work [2], we show that inverse folding can be solved in linear time for a large collection of targets, including every structure that contains no isolated base pair and no isolated stack, i.e. such that $h_{min} = 3$ with h_{min} that denotes the minimum number of base pairs in an helix. This constraint has practical relevance as RNAs are known to contain few helices of small sizes and, even though, such helices are not major contributors to the free-energy/stability.

Our approach introduces a notion of modulo m -separability, generalizing a separability property introduced by Hales *et al.* [3]. (m -)Separability represents a sufficient condition for the existence of a solution to the inverse folding problem. This allows us

to propose a dynamic programming algorithm that, for any input secondary structure of length n , returns a modulo m -separated sequence in time $\mathcal{O}(2^m mn)$ anytime such a sequence exists.

Using this algorithm, we show that every structure with $h_{min} = 3$ is either trivially non-designable (if it contains a forbidden motif) or always admits a modulo-2 separated solution ($m = 2$), which can be obtained in linear time. A trivial extension of our algorithm allows us to uniformly generate them within the set of modulo-2 separable sequences.

Further experiments show that structures featuring shorter helices ($h_{min} = 2$) are often solved with our algorithm with m "small" and that our sequences obtained for $h_{min} = 3$ represent better candidates than random compatible sequences in the more realistic Turner energy model.

References

- [1] Edouard Bonnet, Pawel Rzkazewski, and Florian Sikora. Designing rna secondary structures is hard. *Journal of Computational Biology*, 27(3):302–316, 2020. PMID:32160034.
- [2] Théo Boury, Laurent Bulteau, and Yann Ponty. RNA Inverse Folding Can Be Solved in Linear Time for Structures Without Isolated Stacks or Base Pairs. In Solon P. Pissis and Wing-Kin Sung, editors, *24th International Workshop on Algorithms in Bioinformatics (WABI 2024)*, volume 312, pages 19:1–19:23, Dagstuhl, Germany, 2024.
- [3] Jozef Hales, Alice Héliou, Ján Manuch, Yann Ponty, and Ladislav Stacho. Combinatorial RNA design: Designability and structure-approximating algorithm in watson-crick and nussinov-jacobson energy models. *Algorithmica*, 79(3):835–856, 2017.